MOLECULAR AND CELLULAR BIOLOGY, July 1997, p. 3779–3785 Vol. 17, No. 7
0270-7306/97/$04.00+0

Processing of DNA Prior to Illegitimate Recombination in Mouse Cells

GARY HENDERSON AND J. PAUL SIMONS*

Department of Anatomy and Developmental Biology, Royal Free Hospital School of Medicine, London NW3 2PF, United Kingdom

Received 3 December 1996/Returned for modification 29 January 1997/Accepted 10 April 1997

In mammalian cells, the predominant pathway of chromosomal integration of exogenous DNA is random or illegitimate recombination; integration by homologous recombination is infrequent. Homologous recombination is initiated at double-strand DNA breaks which have been acted on by single-strand exonuclease. To further characterize the relationship between illegitimate and homologous recombination, we have investigated whether illegitimate recombination is also preceded by exonuclease digestion. Heteroduplex DNAs which included strand-specific restriction markers at each of four positions were generated. These DNAs were introduced into mouse embryonic stem cells, and stably transformed clones were isolated and analyzed to determine whether there was any strand bias in the retention of restriction markers with respect to their positions. Some of the mismatches appear to have been resolved by mismatch repair. Very significant strand bias was observed in the retention of restriction markers, and there was polarity of marker retention between adjacent positions. We conclude that DNA is frequently subjected to 5′→3′ exonuclease digestion prior to integration by illegitimate recombination and that the length of DNA removed by exonuclease digestion can be extensive. We also provide evidence which suggests that frequent but less extensive 3′→5′ exonuclease processing also occurs.

Most DNA introduced into mammalian cells is lost; that which is stably maintained is usually retained by virtue of its integration into nonhomologous chromosomal DNA sites by so-called illegitimate recombination. If the exogenous DNA includes sequences which are homologous with chromosomal sequences, a small minority of cells which integrate DNA will do so by homologous recombination. The practical application of homologous recombination between exogenous DNA and a chromosomal target (gene targeting) is making an enormous impact on mouse molecular genetics (4, 5). Highly efficient gene targeting would provide the ideal form of gene therapy, allowing deleterious mutations to be corrected rather than merely compensating for them (42), but at present the efficiency of gene targeting is several orders of magnitude too low for this to be feasible. Improvements in the frequency of homologous recombination and reductions in the proportion of integration which occurs by illegitimate recombination would be required before homologous recombination could be used for gene therapy. For such improvements to be made, it will be important to understand the mechanisms of gene targeting and illegitimate recombination and the relationships between these processes.

There have been numerous studies of the mechanism of homologous recombination in mammalian cells. In many cases, the recombination substrates were introduced into cells together; in these studies, homologous recombination occurs efficiently and prior to chromosomal integration (1, 6, 20). In contrast with gene targeting, such extrachromosomal homologous recombination is efficient. The mechanism of extrachromosomal homologous recombination is different from that of chromosomal homologous recombination and can best be explained by the single-strand annealing model (20). Briefly, double-stranded ends of extrachromosomal recombination substrates are processed by single-strand exonucleases which expose complementary single strands which anneal; this is followed by filling in of single-strand gaps and ligation. In principle, the single-strand annealing pathway of homologous recombination can operate if the polarity of exonuclease digestion of both substrates is either 5′→3′ or 3′→5′. In both yeast cells and *Xenopus* oocytes, 5′→3′ exonuclease degradation has been found to accompany recombination by the single-strand annealing pathway (10, 23, 24).

The best model of chromosomal homologous recombination is the double-strand break (gap) repair model of Szostak et al. (38). In this model, recombination is initiated at a double-strand break or gap in one of the recombination substrates, which is enlarged by 5′→3′ exonuclease. One of the single-stranded 3′ tails invades the duplex of the other substrate and primes DNA synthesis, creating a D loop. The other 3′ single-stranded tail anneals to the displaced single strand and primes DNA synthesis. Ultimately, two Holliday junctions are formed, which are resolved to give recombinant or nonrecombinant products. In yeast, meiotic homologous recombination is initiated at double-strand breaks (36) which are processed by 5′→3′ exonuclease degradation (37), as predicted by the double-strand break (gap) repair model (38). The available evidence indicates that gene targeting by insertion vectors in mammalian cells proceeds by a mechanism broadly similar to that involved in yeast homologous recombination (41). A common feature of the mechanisms of chromosomal and extrachromosomal homologous recombination is the processing of double-stranded DNA ends by single-strand exonuclease digestion.

During gene targeting, nonhomologous DNA is removed from the ends of targeting vectors (17, 22). We have previously shown that during insertion targeting, this is due to exonuclease removal of one or (probably) both strands (17). In contrast, a number of studies found that DNA which has been inte-

* Corresponding author. Mailing address: Department of Anatomy and Developmental Biology, Royal Free Hospital School of Medicine, Rowland Hill St., London NW3 2PF, United Kingdom. Phone: 44 171-794 0500, ext. 4350. Fax: 44 171-830 2917. E-mail: simons@rfhsm.ac.uk.

grated by illegitimate recombination is largely intact, with very little loss of sequence from the ends of the molecules (11, 21, 30, 39, 40). While this might give the impression that there is no removal of DNA from the ends of the molecules prior to integration by illegitimate recombination, it is also possible that the DNA is subjected to single-strand exonuclease degradation followed by resynthesis.

The experiments presented here were designed to investigate whether DNA integrated by illegitimate recombination was processed by exonuclease degradation prior to integration. We prepared heteroduplex DNA in which the two strands differed at a number of restriction sites; these heteroduplexes were introduced into cells, and clones which were obtained were analyzed for the presence or absence of the restriction markers. We have found that a substantial proportion of DNA ends are extensively digested by a 5′→3′ exonuclease prior to integration. The data further suggest that a similar proportion of ends are subjected to 3′→5′ exonuclease digestion, although less extensively.

MATERIALS AND METHODS

Vector construction and preparation of heteroduplexes. The *Bss*HII site in the coding sequence of the neomycin resistance gene (*neo*) of pMC1neoPolA(C) (17) was destroyed while preserving the coding potential, by oligonucleotide-directed mutagenesis using the Altered Sites II system (Promega). The *Xho*I-*Sal*I *neo* cassette was blunted and inserted into the blunted *Bam*HI site of pBluescriptII KS(+) (Stratagene) which had previously had its *Eco*RI site destroyed by filling in and recircularization. The 1,777-bp *Xba*I-*Apa*I fragment of the sheep β-lactoglobulin gene (GenBank accession no. X12817) was blunted and inserted into the *Eco*RV site in the pBluescriptII KS(+) polylinker, and the 1,884-bp *Nco*I β-lactoglobulin fragment was filled in and inserted into the filled in *Xba*I site of the polylinker. The construct was linearized at the *Cla*I site in the polylinker (position D), and the ends were partially filled in with the Klenow fragment of DNA polymerase I and dCTP; linkers which include *Eco*RI (GCG GAATTCCG) or *Sac*I (GCGGAGCTCCG) sites were ligated into the filled in *Cla*I site. The resulting constructs were linearized at the *Bcl*I site in the *Xba*I-*Apa*I β-lactoglobulin fragment (position C) and partially filled in with dGTP before insertion of *Pvu*II (ATCGCCAGCTGGC) or *Nde*I (ATCGCCATATG GC) linkers. *Bam*HI (CGGGATCCCG) or *Kpn*I (CGGGTACCCG) linkers were inserted into the *Nru*I site in the *Nco*I β-lactoglobulin fragment (position B), and *Bcl*I (CCGGTGATCACC) or *Bsr*GI (CCGGTGTACACC) linkers were inserted into the *Not*I site in the polylinker (position A), after filling in of the vector ends with dGTP. All linker insertions were verified by sequencing. The two vectors obtained, designated P6 and N3, have the *Bcl*I, *Bam*HI, *Pvu*II, and *Eco*RI linkers and the *Bsr*GI, *Kpn*I, *Nde*I, and *Sac*I linkers, respectively; these vectors give the *neo* antisense strand after single-strand rescue. P6 and N3 were digested with *Bss*HII to excise the entire inserts (including polylinkers) and recloned into *Bss*HII-digested pBluescriptII KS(+) to obtain the identical sequences in inverted orientation (P2 and N2, respectively), which give the *neo* sense strand after single-strand rescue.

Single-stranded DNA from each construct was rescued following infection with R408 helper phage and purified by the methods recommended by Stratagene. To excise the inserts from single-stranded N3 and P6, the regions around the *Bss*HII sites were made double stranded by annealing the circular single-stranded DNA with 18-nucleotide oligonucleotides complementary to the restriction sites and neighboring sequences, followed by digestion with *Bss*HII. Circular single-stranded DNA was mixed with oligonucleotides TTAATTGCG CGCTTGGCG and CAGTGAGCGCGCGTAATA in 50 μl of *Bss*HII reaction buffer (100 mM NaCl, 10 mM bis Tris propane-HCl, 10 mM $MgCl_2$, 1 mM dithiothreitol [pH 7.0] at 25°C), heated to 75°C for 5 min, and allowed to cool slowly to room temperature; the annealed DNA was digested with *Bss*HII at 50°C. The same procedure was adopted for clones N2 and P2 except that oligonucleotides TATTACGCGCGCTTGGCG and CAGTGAGCGCGCAAT TAA were used. To generate the heteroduplex and homoduplex DNAs, the digested single-stranded DNAs were mixed in equimolar amounts in the appropriate combinations, heated to 75°C, and allowed to cool slowly to room temperature. The annealed double-stranded DNA was gel purified by using Geneclean (Bio 101), quantitated, and used for electroporation of embryonic stem (ES) cells.

Cell culture. The mouse ES cell line E14 (13) was used throughout these experiments, cultured as previously described (17) except that 10% ES-qualified fetal calf serum (Life Technologies) was used. Cells at 10^8 ml^{-1} in phosphate-buffered saline were mixed with DNA (0.3 μg ml^{-1}) and electroporated at 800 V (2,000 V cm^{-1}) and 3 μF. After 10 min at room temperature, the cells were dispersed in complete medium at 10^6 ml^{-1} and plated in 6-cm-diameter petri dishes (5×10^6 per dish). Selection (0.3 mg of Geneticin ml^{-1}) was applied the following day, and colonies were picked after 10 to 13 days of selection and expanded for DNA preparation.

DNA analysis. DNA was isolated from confluent 24-well plates by the method of Laird et al. (18) or by using a Puregene kit. The patterns of retention of restriction markers were analyzed by PCR amplification of short DNA segments flanking each marker position, followed by restriction digestion. PCR products were digested under the recommended conditions except that the PCR products were included in the digests without removal of any of the components of the PCRs.

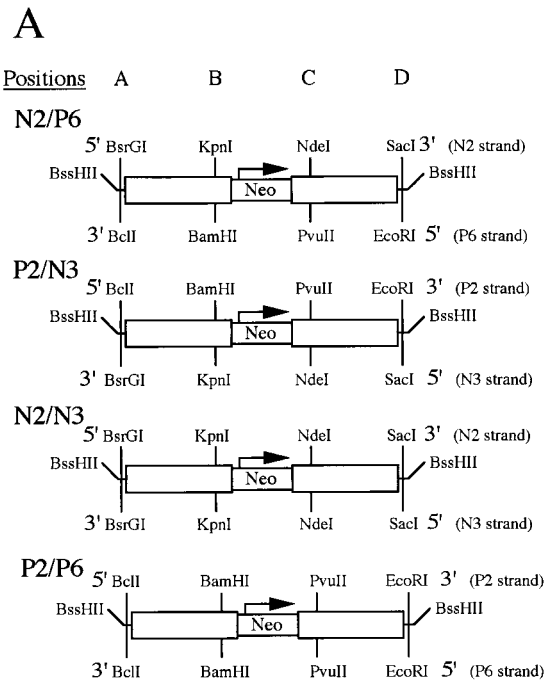
Statistical analysis. Statistical analysis was performed as described by Siegel (35), using the χ^2 test; for the analysis of 2×2 contingency tables, Yates' correction for continuity was applied.

RESULTS

Experimental design. To test the hypothesis that DNA is subjected to frequent single-stranded exonucleolytic degradation prior to nontargeted integration, we used heteroduplex DNA molecules which included the selectable *neo* gene and in which the two strands were distinguished by restriction site differences.

Constructs were built such that *neo* was placed centrally, with 1.8 and 1.9 kb of stuffer DNA on either side, into which the restriction markers were placed. In this way, we ensured that *neo* was perfectly matched in the heteroduplexes and that the data would not be biased because of damage to the gene. The restriction markers were placed at four positions in the constructs, two on either side of *neo*: constructs which had one set of restriction markers or another (P and N sets) were built; the restriction markers were chosen such that at each position, the differences between the markers would be in the central dinucleotide of each restriction recognition sequence. On either side of *neo*, one marker was placed close to *neo* (1.4 kb from the nearest end), and one lay very close (<100 bp) to the end of the fragment after excision from the vector backbone. The excised fragments were roughly symmetrical with respect to the positions of the polymorphic restriction sites. Single-stranded DNA was prepared from these clones, and the inserts were excised and annealed to give heteroduplexes which have mismatches at each of the four restriction sites. To control for any strand-specific effects, the two reciprocal heteroduplexes (N2/P6 and P2/N3) were prepared; the two homoduplexes (N2/N3 and P2/P6) were generated by the same procedure and used as controls; double-stranded insert from plasmid N2 was used as a further control. The structures of the heteroduplexes and homoduplexes are depicted in Fig. 1. Each of these DNAs was separately introduced into mouse ES cells.

Following electroporation of the heteroduplexes and selection for G418 resistance, colonies were picked and expanded. DNA from each clone was amplified by PCR, using primers which flank each of the four polymorphic restriction sites, and digested with the appropriate restriction enzymes. Following electroporation with heteroduplex DNA, 58 and 90 clones were derived (N2/P6 and P2/N3 heteroduplexes, respectively); 8 and 6 clones were derived from the homoduplexes (N2/N3 and P2/P6, respectively), and 32 clones were derived from the N2 double-stranded insert. In the majority of clones, it appears that there was loss of terminal sequences prior to integration: from most clones, we consistently failed to obtain a specific PCR product from one or both ends (Table 1). This was the case with the original primer sets, in which the 5′ ends of the most terminal primers were 0 and 8 bp from the ends of molecules which were electroporated into cells; almost identical results were obtained with primers of which the 5′ ends were 18 and 22 bp from the ends. Only 18 clones (~9%) appeared to have integrated full-length construct. The frequent loss of terminal sequences was common to heteroduplex DNA-derived clones, the clones derived from the control homoduplexes N2/N3 and P2/P6, and those derived from the



B

Positions A B C D

```
N2  5'..TA.....TA......TA.....GC..3'
P6  3'..TA.....TA......CG.....TA..5'

P2  5'..AT.....AT.....GC.....AT..3'
N3  3'..AT.....AT.....AT.....CG..5'
```

FIG. 1. DNAs electroporated into cells. (A) Gross structures of the heteroduplexes and homoduplex DNAs which were introduced into ES cells. The double-stranded DNA excised from plasmid N2 has the same structure as the N2/N3 homoduplex except that both strands were derived from N2. Upstream of *neo*, the *Bcl*I and *Bsr*GI restriction markers (position A) are 59 bp from the ends and the *Bam*HI and *Kpn*I markers (position B) are 1.4 kb from the ends; downstream of *neo*, the *Pvu*II and *Nde*I markers (position C) are 1.4 kb from the ends and the *Eco*RI and *Sac*I markers (position D) are 72 bp from the ends (distances are given from the closest end). The empty boxes represent the stuffer fragments from the sheep β-lactoglobulin gene. The arrows indicate the direction of *neo* transcription. (B) Mismatches incorporated into the heteroduplex DNAs.

double-stranded N2 insert, indicating that the loss of terminal sequences was not caused by the mismatches. Similarly, the frequencies of loss of internal markers were similar for the clones obtained with each of the DNAs (Table 1).

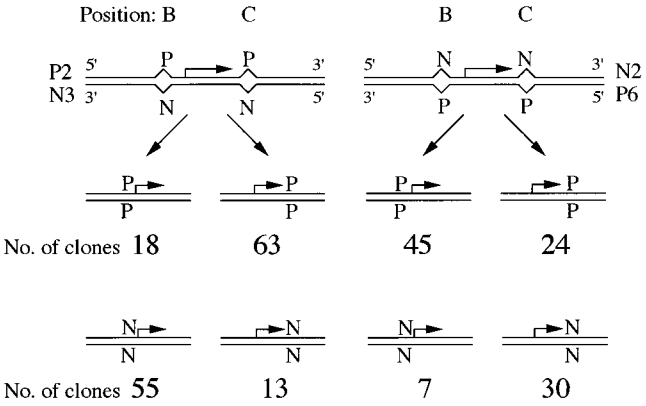


FIG. 2. Retention of markers at positions B and C. The heteroduplexes are represented in the top line; the polarities of the DNA strands are shown, and the mismatches at positions B and C only are indicated. The lower two lines show the structures and the numbers of clones which have retained N- or P-strand-derived markers at positions B and C. The arrows indicate the direction of transcription of *neo*.

In the vast majority of cases (91%), the PCR products were cut to completion by one restriction enzyme and were not cut by the other. In each of the clones in which PCR products were cut by both enzymes, other positions clearly possessed one marker or the other. At one position of each of two clones, neither restriction enzyme cut the PCR products. Southern blot analysis of 34 representative clones indicated that in each case, a single copy of the DNA was integrated (data not shown). Together, these findings are entirely consistent with single copies of the DNA being integrated into the host cell's genome in the majority of cases. A number of conclusions can be drawn from the frequencies with which one or other of the restriction enzyme recognition sequences is retained along with the patterns of retention of sequences at different positions. Because of the predominant loss of terminal sequences, the sample sizes for the internal positions (B and C) are greatest, and these data are considered first.

Nonrandom loss of markers at internal positions. Figure 2 summarizes the data from positions B and C, considered independently of any other positions. The numbers of clones which retain the P-strand and the N-strand restriction markers are shown for each position and for each heteroduplex. For example, taking position B and heteroduplex P2/N3, 18 clones retained the P-strand marker and 55 retained the N-strand marker; this is a significant deviation from the frequencies expected if there was no bias in the loss or retention of markers from the different strands ($P < 0.001$). At position C of heteroduplex P2/N3, the P-strand marker was preferentially retained ($P < 0.001$), as was the P-strand marker at position B of

TABLE 1. Frequencies of retention of DNA at each of the marked positions

DNA	No. of clones which retained sequences[a] (%)				Total no. of clones
	Position A	Position B	Position C	Position D	
Heteroduplex N2/P6	19 (33)	56 (97)	58 (100)	14 (24)	58
Heteroduplex P2/N3	27 (30)	81 (90)	79 (88)	24 (27)	90
Homoduplex N2/N3	0 (0)	6[b] (86)	7[b] (100)	2 (25)	8
Homoduplex P2/P6	1 (17)	5 (83)	5 (83)	2 (33)	6
N2 double-stranded insert	9 (28)	26 (81)	26 (81)	8 (25)	32

[a] Number of clones which gave a product of the expected size following PCR of genomic DNA.
[b] Seven clones analyzed.

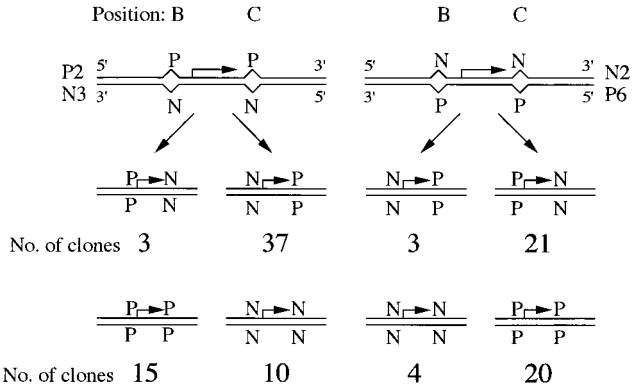


FIG. 3. Polarity in the retention of markers at positions B and C. The heteroduplexes are represented in the top line; the polarities of the DNA strands are shown, and the mismatches at positions B and C only are indicated. The arrows indicate the direction of transcription of *neo*.

heteroduplex N2/P6 ($P < 0.001$). At position C of heteroduplex N2/P6, there was no significant strand bias ($0.3 < P < 0.5$).

Analysis of both heteroduplexes together allows us to dissect the results for effects which are related to the positions and effects related to the sequences of the two strands. This analysis reveals that the markers closest to the 5′ end of each strand are preferentially lost for each of positions B and C ($P < 0.001$ for both). The analysis further shows that there is no strand specificity for P- or N-strand markers at position B ($0.1 < P < 0.2$) but that at position C, there is a significant preference for the retention of the P-strand marker ($P < 0.005$). The preferential loss of markers closest to the 5′ ends is consistent with 5′→3′ exonuclease digestion; the P-strand preference at position C is discussed below.

If 5′→3′ exonuclease does digest transforming DNAs, it is predicted that there should be polarity in the loss or retention of sequence: in individual DNA molecules, more 3′ markers are predicted to be retained more frequently than 5′ markers. The informative clones in this respect are those which have retained markers which were derived from different strands of the heteroduplexes at two positions. The data on the patterns of retention of the P- and N-strand markers at positions B and C are shown in Fig. 3. In those clones which have retained sequences from different strands at positions B and C, the sequences closest to the 3′ end of each strand are retained preferentially over the sequences closest to the 5′ end of each strand. The preference for retention of 3′ markers is significant for each heteroduplex ($P < 0.001$ for both), as expected from the analysis of positions B and C independently. Analysis of the combined data for the two reciprocal heteroduplexes confirms that the bias in retention of markers is due to the polarity of the DNA ($P < 0.001$) and not because of sequence effects ($0.8 < P < 0.9$).

These results are consistent with the prediction that the heteroduplexes were processed by 5′→3′ exonuclease. The degree of polarity is almost exactly that predicted from the analysis of positions B and C independently ($0.5 < P < 0.7$ for each heteroduplex).

No strand preference in the maintenance of terminal markers. In most clones, we reproducibly failed to amplify the sequences from positions A and D. Because of the small sample size, the data from both ends of both heteroduplexes have been analyzed together for any strand preference effects. In the clones which had incorporated the terminal sequences, there was no evident preference for maintenance of one strand or the other (Table 2): in 32 cases the 5′-end-derived markers were retained, whereas 37 clones retained 3′-end-derived sequences ($0.5 < P < 0.7$). The data have been examined for any evidence of polarity between the terminal markers and the adjacent internal markers (i.e., between positions A and B and between positions C and D together). Although there was no preferential retention of either strand at positions A and D, there is evidence for polarity between these markers and internal markers (Table 3), with markers close to the 5′ end of the DNA strand being retained preferentially (16:5, $P < 0.02$). This is the opposite of the polarity predicted if the heteroduplex DNAs had been subjected to 5′→3′ exonuclease digestion. These data are clearly at odds with those from internal positions and suggest the existence of different pathways in the processing of the transforming DNAs.

TABLE 2. Frequencies of retention of markers at terminal positions

Marker retained	No. of clones	
	Position A	Position D
Heteroduplex N2/P6		
N	6[a]	7[b]
P	8[b]	6[a]
Heteroduplex P2/N3		
N	8[b]	8[a]
P	12[a]	14[b]

[a] In the heteroduplex, the retained marker was closer to the 5′ end of the strand.
[b] In the heteroduplex, the retained marker was closer to the 3′ end of the strand.

TABLE 3. Polarity of retention of markers between terminal and adjacent internal positions

Marker retained	No. of clones	
	Positions A and B	Positions D and C
Heteroduplex N2/P6		
5′5′[a]	3[e]	2[h]
5′3′[b]	3[f]	2[i]
3′5′[c]	0[g]	3[j]
3′3′[d]	7[h]	3[e]
Heteroduplex P2/N3		
5′5′[a]	2[h]	4[e]
5′3′[b]	7[g]	4[j]
3′5′[c]	2[f]	0[i]
3′3′[d]	5[e]	11[h]

[a] In the heteroduplex, both the retained terminal marker and the retained internal marker (position B or C) were closer to the 5′ end of the strand.
[b] In the heteroduplex, the retained terminal marker (position A or D) was closer to the 5′ end of the strand, and the retained internal marker (position B or C) was closer to the 3′ end of the strand.
[c] In the heteroduplex, the retained terminal marker (position A or D) was closer to the 3′ end of the strand, and the retained internal marker (position B or C) was closer to the 5′ end of the strand.
[d] In the heteroduplex, both the retained terminal marker and the retained internal marker (position B or C) were closer to the 3′ end of the strand.
[e] N-strand marker at both positions.
[f] N-strand marker at position A and P-strand marker at position B.
[g] P-strand marker at position A and N-strand marker at position B.
[h] P-strand marker at both positions.
[i] N-strand marker at position C and P-strand marker at position D.
[j] P-strand marker at position C and N-strand marker at position D.

TABLE 4. Frequencies of retention of markers at internal positions in subsets of clones which have lost or retained the adjacent terminal markers

Marker retained	No. of clones			
	Flanking marker lost		Flanking marker retained	
	Position B	Position C	Position B	Position C
Heteroduplex N2/P6				
N	3[a]	23[b]	4[a]	7[b]
P	31[b]	19[a]	14[b]	5[a]
Heteroduplex P2/N3				
N	42[b]	8[a]	13[b]	5[a]
P	9[a]	47[b]	9[a]	16[b]

[a] In the heteroduplex, the retained marker was closer to the 5′ end of the strand.
[b] In the heteroduplex, the retained marker was closer to the 3′ end of the strand.

If different pathways of DNA processing exist, we would predict that the internal positions would behave differently when the external markers are retained or are lost. The data are summarized in Table 4. At position B, there is a significant difference between the subsets in this respect ($P < 0.05$). Despite this, there is a significant preference for retention of 3′-end-derived markers both in the subset of clones which have lost the sequences at position A ($P < 0.001$) and in those which have retained sequences at position A ($P < 0.05$). The magnitude of the excess of 3′-end-derived markers is greater in the subset of clones which lost the sequences at position A. At position C, there is no significant difference between the two subsets of data ($0.9 < P < 0.95$). At this position, there is a significant preference for retention of 3′-end-derived markers in those clones which lost the sequences at position D ($P < 0.001$) but not in those which retained sequences at position D ($0.05 < P < 0.1$). The previously identified bias toward retention of the P-strand marker at position C is expected to reduce the sensitivity of these comparisons and may explain the differences between the results for positions B and C.

DISCUSSION

There are three broad mechanisms which could be involved in the removal of mismatches in heteroduplex DNA: repair, 5′→3′ exonuclease digestion, and 3′→5′ exonuclease digestion. Additionally, it is possible for heteroduplexes to remain unchanged until replicated. Mismatches which are resolved by replication after DNA integration would give rise to a mixed population of cells with one or other marker. From the data we present, it is clear that there is very significant nonrandomness in the loss and retention of sequences following integration into the ES cell genome. The use of heteroduplexes prepared from defined single strands and repeating the experiments with the reciprocal heteroduplexes were critical elements of the experimental design. The data allow effects to be attributed unequivocally to the locations of the sequences with respect to the 5′ and 3′ ends of the DNA strand (independent of the nature of the sequence) and, in one case, to the nature of the sequences (independent of the location). The data suggest multiple pathways of heteroduplex DNA processing, including mismatch repair and 5′→3′ and 3′→5′ exonuclease degradation.

DNA repair. Mismatched bases on complementary DNA strands normally arise in a number of ways, and mechanisms exist for the correction of such mismatches (28, 29), the predominant mechanism being known as mismatch repair. Mismatch repair in eukaryotes appears to be performed by a pathway similar to that in *Escherichia coli* (8). Following the detection of a mismatch, repair is initiated by a single-strand nick which may be 5′ or 3′ of the mismatch. In both *E. coli* and human cells, the single-strand nick is expanded to a gap by either 5′→3′ or 3′→5′ exonuclease; in human cells, the excision tract terminates 90 to 170 nucleotides beyond the position of the mismatch (8) and the gap is filled by resynthesis. In mammalian cells, G-T mismatches (which are continually generated by deamination of 5-methyl cytosine) are efficiently repaired with a strong preference for repair to G-C (2). Other mispairs have been found to be repaired with variable efficiencies and, when the mispair includes a G or a C, with bias toward G-C (3).

The preferential loss of markers closer to the 5′ ends of the DNA molecules at internal positions cannot be explained on the basis of mismatch repair. Despite this, the evidence suggests that mismatch repair did occur in some of the clones. In 9 of 49 informative clones, a marker for one strand is flanked by markers derived from the other strand (P-N-P or N-P-N). These patterns cannot be explained solely by exonuclease degradation or replication. The simplest explanation of these patterns is that the central mismatch has been repaired in these clones. In addition, the preferential retention of P-strand markers at position C is consistent with preferential mismatch repair.

At position C, the mismatches were

5′--GC--3′
| |
3′--AT--5′

with the upper line representing the P strand. Brown and Jiricny (3) found that G-A mismatches are repaired to G-C about twice as frequently as to T-A and that C-T is repaired to C-G fivefold more often than to A-T. If correction of dinucleotide mismatches has specificity similar to that of repair of single-base mismatches, mismatch repair would strongly favor correction to the P-strand sequence at position C. Thus, the most likely explanation of the preferential retention of the P-strand marker at position C is directional mismatch repair. The P-strand preference at this position was superimposed on another source of nonrandomness (see below), resulting in exaggerated preference for the P-strand marker at this position in the P2/N3 heteroduplex and no apparent preference in the N2/P6 heteroduplex.

From the data for the other three positions, there is no evidence for the occurrence of directional mismatch repair. At both positions A and B, the mismatches were

5′--AT--3′
| |
3′--AT--5′

and so directional mismatch repair would not be expected. At position D, the mismatches were

5′--AT--3′
| |
3′--CG--5′

which does have the potential for directional mismatch repair, favoring the N (lower) strand; there was no evidence for preferential retention of N-strand sequences. Position D is very close to the end of the DNA molecule, and it may be that the

mismatched bases were rapidly removed by exonuclease digestion (see below).

In contrast to mismatch repair, excision repair (14) is responsible for the elimination of genetic damage such as T-T dimers, abasic sites, and DNA adducts. Excision repair cannot account for any of the nonrandomness that we have observed in the processing of heteroduplex DNAs: resolution of mismatches by this pathway is very inefficient (15), strand-specific differences in excision repair have been found only within transcription units (25), and the mechanism by which strand-specific excision repair is initiated (7) would not operate on mismatched bases.

5′→3′ exonuclease digestion of transforming DNA. There are two lines of evidence which show that the transforming DNA is subjected to 5′→3′ exonuclease degradation: the excessive loss of the markers at internal positions which are closer to the 5′ ends of the DNA strands, and the polarity between the internal positions in the loss of markers. The amount of DNA removed by exonuclease digestion must be extensive because the restriction markers at positions B and C, with which the 5′→3′ exonuclease degradation was detected, are 1.4 kb from the ends of the DNA molecules. There are, however, clear limits to the extent of exonuclease degradation which is compatible with the data. In an extreme case, exonuclease digestion would degrade a complete strand, leaving the other strand intact. While single-stranded DNA has been shown to be competent for stable transformation (31), if one or other strand were completely removed, no strand preference or polarity would have been observed. While it is not possible to estimate from the data what proportion of ends are processed by 5′→3′ exonuclease digestion, it is clear that a substantial fraction of (perhaps most) DNA molecules are processed in this manner. Presumably, given that exonuclease digestion was frequent, a proportion of DNA molecules were subjected to degradation at both ends.

Additional pathways of DNA processing. The predominant loss of terminal sequences observed is at odds with previous data which indicate that terminal sequences are generally preserved with little or no loss (11, 21, 32, 39, 40). In the experiments reported here, if more than about 20 bp of DNA were lost from the ends of the integrated DNA, we would fail to amplify the terminal markers by PCR, and it is possible that this is the explanation of the discrepancy. The finding that the heteroduplexes, the control homoduplexes, and the control double-stranded insert all behave similarly in this respect shows that this loss of sequences is not an artifact of the use of mismatched DNAs or of the method of preparation of the heteroduplexes.

Where the terminal sequences were retained, the data for the terminal positions alone provide no evidence for exonuclease digestion. There are a number of ways in which this observation can be explained. First, rapid integration may have occurred, such that there was no opportunity for exonuclease processing of the DNA before its incorporation into chromosomal DNA. Alternatively, the DNA ends may have been protected from exonuclease action. In this respect, it is interesting that there are at least two abundant nuclear proteins which have DNA end-binding activity: both poly(ADP-ribose) polymerase and Ku protein bind to double-stranded DNA ends (27, 34) and protect bound DNA ends from nucleases (12, 27, 34). Both poly(ADP-ribose) polymerase (9, 43) and Ku protein (16) have been reported to be involved in nonhomologous recombination. Either or both of these proteins could protect DNA ends from exonuclease attack prior to chromosomal integration.

While we cannot exclude the possibility that a proportion of DNA ends are protected from exonuclease attack by such proteins, the data are inconsistent with such protection of all DNA ends in the subsets of clones in which the terminal positions are retained. One further possibility to explain the lack of any apparent strand bias at terminal positions is that 3′→5′ exonuclease digestion occurs at a frequency which is similar to that of 5′→3′ exonuclease digestion. The polarity observed between the terminal markers and the internal markers supports this hypothesis. To be consistent with the data, there would have to be significant differences in the lengths of the tracts degraded by the 5′→3′ and 3′→5′ exonucleases. A major difference between the products of 5′→3′ exonuclease and of 3′→5′ exonuclease digestion is that 3′→5′ exonuclease-digested DNA leaves a 5′ single-stranded overhang which could be filled in by resynthesis prior to chromosomal integration. Thus, it may be that shorter regions of DNA are lost by 3′→5′ exonuclease digestion because the removal of DNA is balanced by resynthesis. The data from clones which have retained terminal sequences indicate (i) no strand preference at terminal positions, (ii) processing by 5′→3′ exonuclease (excess of 3′-end-derived markers at position B), and (iii) processing by 3′→5′ exonuclease (polarity between the terminal and adjacent markers). These data can be reconciled if (in at least some cases) 3′→5′ exonuclease digestion and resynthesis preceded more extensive 5′→3′ exonuclease digestion.

In the majority of clones, the terminal sequences were lost. Possible mechanisms by which the loss occurred are double-strand endonuclease cleavage, 5′→3′ exonuclease digestion past a nick in the other strand, and exonuclease digestion of both strands. Given the evidence for processing by both 5′→3′ and 3′→5′ exonucleases, the loss of terminal DNA sequences prior to integration may be due to exonuclease degradation of both strands, although other mechanisms cannot be excluded.

DNA processing and illegitimate and homologous recombination. Results from a number of systems indicate that chromosomal and extrachromosomal homologous recombination are initiated by 5′→3′ exonuclease degradation of DNA at a double-strand break. The data presented here extend these findings to show that prior to illegitimate recombination, DNA is frequently subjected to 5′→3′ exonuclease digestion. Examination of the sequences of DNAs joined by illegitimate recombination has shown that frequently there are short regions of homology between the parent DNA molecules at the junction points (19, 26, 33). Illegitimate recombination in such cases probably proceeds by a mechanism analogous to the single-strand annealing pathway of homologous recombination but in which exposed single strands anneal at regions of very limited homology (19, 26, 30). The data presented here suggest that illegitimate recombination mediated by single-strand annealing of short homologies follows 5′→3′ exonuclease digestion.

Failure of a cell to repair double-strand breaks in chromosomal DNA would in most cases be expected to have detrimental or lethal effects. To preserve the integrity of the genome, repair of double-strand breaks by homologous recombination (e.g., sister chromatid exchange) would be the preferred pathway. It is likely that illegitimate recombination provides an error-prone fallback mechanism which operates for cells which fail to repair double-strand breaks by homologous recombination (e.g., for cells in G_1, which cannot undergo sister chromatid exchange). In this context, it is perhaps to be expected that the initial steps in processing of double-strand DNA ends should be common to both illegitimate and homologous recombination.

ACKNOWLEDGMENTS

We thank Raya Al-Shawi and Satish Kumar for helpful comments on the manuscript.

This work was supported by the U.K. Medical Research Council (grant G9404715MB).

REFERENCES

1. **Bollag, R. J., A. S. Waldman, and R. M. Liskay.** 1989. Homologous recombination in mammalian cells. Annu. Rev. Genet. **23:**199–225.
2. **Brown, T. C., and J. Jiricny.** 1987. A specific mismatch repair event protects mammalian cells from loss of 5-methylcytosine. Cell **50:**945–950.
3. **Brown, T. C., and J. Jiricny.** 1988. Different base/base mispairs are corrected with different efficiencies and specificities in monkey kidney cells. Cell **54:** 705–711.
4. **Capecchi, M. R.** 1989. Altering the genome by homologous recombination. Science **244:**1288–1292.
5. **Capecchi, M. R.** 1989. The new mouse genetics: altering the genome by gene targeting. Trends Genet. **5:**70–76.
6. **Carroll, D., S. H. Wright, R. K. Wolff, E. Grzesiuk, and E. B. Maryon.** 1986. Efficient homologous recombination of linear DNA substrates after injection into *Xenopus laevis* oocytes. Mol. Cell. Biol. **6:**2053–2061.
7. **Donahue, B. A., S. Yin, J. S. Taylor, D. Reines, and P. C. Hanawalt.** 1994. Transcript cleavage by RNA polymerase II arrested by a cyclobutane pyrimidine dimer in the DNA template. Proc. Natl. Acad. Sci. USA **91:**8502–8506.
8. **Fang, W., and P. Modrich.** 1993. Human strand-specific mismatch repair occurs by a bidirectional mechanism similar to that of the bacterial reaction. J. Biol. Chem. **268:**11838–11844.
9. **Farzaneh, F., G. N. Panayotou, L. D. Bowler, B. D. Hardas, T. Broom, C. Walther, and S. Shall.** 1988. ADP-ribosylation is involved in the integration of foreign DNA into the mammalian cell genome. Nucleic Acids Res. **16:** 11319–11326.
10. **Fishman-Lobell, J., N. Rudin, and J. E. Haber.** 1992. Two alternative pathways of double-strand break repair that are kinetically separable and independently modulated. Mol. Cell. Biol. **12:**1292–1303.
11. **Folger, K. R., E. A. Wong, G. Wahl, and M. R. Capecchi.** 1982. Patterns of integration of DNA microinjected into cultured mammalian cells: evidence for homologous recombination between injected plasmid DNA molecules. Mol. Cell. Biol. **2:**1372–1387.
12. **Getts, R. C., and T. D. Stamato.** 1994. Absence of a Ku-like DNA end binding activity in the *xrs* double-strand DNA repair-deficient mutant. J. Biol. Chem. **269:**15981–15984.
13. **Handyside, A. H., G. T. O'Neill, M. Jones, and M. L. Hooper.** 1989. Use of BRL-conditioned medium in combination with feeder layers to isolate a diploid embryonal stem-cell line. Rouxs Arch. Dev. Biol. **198:**48–55.
14. **Huang, J. C., D. L. Svoboda, J. T. Reardon, and A. Sancar.** 1992. Human nucleotide excision nuclease removes thymine dimers from DNA by incising the 22nd phosphodiester bond 5′ and the 6th phosphodiester bond 3′ to the photodimer. Proc. Natl. Acad. Sci. USA **89:**3664–3668.
15. **Huang, J. C., D. S. Hsu, A. Kazantsev, and A. Sancar.** 1994. Substrate spectrum of human excinuclease, repair of abasic sites, methylated bases, mismatches, and bulky adducts. Proc. Natl. Acad. Sci. USA **91:**12213–12217.
16. **Jeggo, P. A., and J. Smith-Ravin.** 1989. Decreased stable transfection frequencies of six X-ray-sensitive CHO strains, all members of the xrs complementation group. Mutat. Res. **218:**75–86.
17. **Kumar, S., and J. P. Simons.** 1993. The effects of terminal heterologies on gene targeting by insertion vectors in embryonic stem cells. Nucleic Acids Res. **21:**1541–1548.
18. **Laird, P. W., A. Zijderveld, K. Linders, M. A. Rudnicki, R. Jaenisch, and A. Berns.** 1991. Simplified mammalian DNA isolation procedure. Nucleic Acids Res. **19:**4293.
19. **Lehman, C. W., J. K. Trautman, and D. Carroll.** 1994. Illegitimate recombination in Xenopus: characterization of end-joined junctions. Nucleic Acids Res. **22:**434–442.
20. **Lin, F.-L., K. Sperle, and N. Sternberg.** 1984. Model for homologous recombination during transfer of DNA into mouse cells: role for DNA ends in the recombination process. Mol. Cell. Biol. **4:**1020–1034.
21. **Macleod, D., R. Lovell-Badge, S. Jones, and I. Jackson.** 1991. A promoter trap in embryonic stem (ES) cells selects for integration of DNA into CpG islands. Nucleic Acids Res. **19:**17–23.
22. **Mansour, S. L., K. R. Thomas, and M. R. Capecchi.** 1988. Disruption of the proto-oncogene int-2 in mouse embryo-derived stem cells: a general strategy for targeting mutations to non-selectable genes. Nature **336:**348–352.
23. **Maryon, E. B., and D. Carroll.** 1989. Degradation of linear DNA by a strand-specific exonuclease activity in *Xenopus laevis* oocytes. Mol. Cell. Biol. **9:**4862–4871.
24. **Maryon, E. B., and D. Carroll.** 1991. Involvement of single-stranded tails in homologous recombination of DNA injected into *Xenopus laevis* oocyte nuclei. Mol. Cell. Biol. **11:**3268–3277.
25. **Mellon, I., G. Spivak, and P. C. Hanawalt.** 1987. Selective removal of transcription-blocking DNA damage from the transcribed strand of the mammalian DHFR gene. Cell **51:**241–249.
26. **Merrihew, R. V., K. Marburger, S. L. Pennington, D. B. Roth, and J. H. Wilson.** 1996. High-frequency illegitimate integration of transfected DNA at preintegrated target sites in a mammalian genome. Mol. Cell. Biol. **16:**10–18.
27. **Mimori, T., and J. A. Hardin.** 1986. Mechanism of interaction between Ku protein and DNA. J. Biol. Chem. **261:**10375–10379.
28. **Modrich, P.** 1987. DNA mismatch correction. Annu. Rev. Biochem. **56:**435–466.
29. **Modrich, P.** 1991. Mechanisms and biological effects of mismatch repair. Annu. Rev. Genet. **25:**229–253.
30. **Nicolás, A. L., P. L. Munz, and C. S. H. Young.** 1995. A modified single-strand annealing model best explains the joining of DNA double-strand breaks in mammalian cells and cell extracts. Nucleic Acids Res. **23:**1036–1043.
31. **Rauth, S., K. Y. Song, D. Ayares, L. Wallace, P. D. Moore, and R. Kucherlapati.** 1986. Transfection and homologous recombination involving single-stranded DNA substrates in mammalian cells and nuclear extracts. Proc. Natl. Acad. Sci. USA **83:**5587–5591.
32. **Roth, D. B., T. N. Porter, and J. H. Wilson.** 1985. Mechanisms of nonhomologous recombination in mammalian cells. Mol. Cell. Biol. **5:**2599–2607.
33. **Roth, D. B., and J. H. Wilson.** 1986. Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. Mol. Cell. Biol. **6:**4295–4303.
34. **Sastry, S. S., K. G. Buki, and E. Kun.** 1989. Binding of adenosine diphosphoribosyltransferase to the termini and internal regions of linear DNAs. Biochemistry **28:**5670–5680.
35. **Siegel, S.** 1956. Nonparametric statistics for the behavioral sciences. McGraw-Hill Kogakusha Ltd., Tokyo, Japan.
36. **Sun, H., D. Treco, N. P. Schultes, and J. W. Szostak.** 1989. Double strand breaks at an initiation site for meiotic gene conversion. Nature **338:**87–90.
37. **Sun, H., D. Treco, and J. W. Szostak.** 1991. Extensive 3′-overhanging, single-stranded DNA associated with the meiosis-specific double-strand breaks at the ARG4 recombination initiation site. Cell **64:**1155–1161.
38. **Szostak, J. W., T. L. Orr-Weaver, R. J. Rothstein, and F. W. Stahl.** 1983. The double-strand-break repair model for recombination. Cell **33:**25–35.
39. **Thomas, K. R., K. R. Folger, and M. R. Capecchi.** 1986. High frequency targeting of genes to specific sites in the mammalian genome. Cell **44:**419–428.
40. **Toneguzzo, F., A. Keating, S. Glynn, and K. McDonald.** 1988. Electric field-mediated gene transfer: characterization of DNA transfer and patterns of integration in lymphoid cells. Nucleic Acids Res. **16:**5515–5532.
41. **Valancius, V., and O. Smithies.** 1991. Double-strand gap repair in a mammalian gene-targeting reaction. Mol. Cell. Biol. **11:**4389–4397.
42. **Vega, M. A.** 1991. Prospects for homologous recombination in human gene therapy. Hum. Genet. **87:**245–253.
43. **Waldman, B. C., and A. S. Waldman.** 1990. Illegitimate and homologous recombination in mammalian cells, differential sensitivity to an inhibitor of polyADP-ribosylation. Nucleic Acids Res. **18:**5981–5988.